

**University of Groningen**

**Building and testing age models for radiocarbon dates in Lateglacial and Early Holocene sediments**

Blockley, S. P. E.; Ramsey, C. Bronk; van der Plicht, J.; Blaauw, Maarten

*Published in:*  
Quaternary Science Reviews

*DOI:*  
[10.1016/j.quascirev.2007.06.007](https://doi.org/10.1016/j.quascirev.2007.06.007)

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*  
Publisher's PDF, also known as Version of record

*Publication date:*  
2007

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

Blockley, S. P. E., Ramsey, C. B., van der Plicht, J., & Blaauw, M. (2007). Building and testing age models for radiocarbon dates in Lateglacial and Early Holocene sediments. *Quaternary Science Reviews*, 26(15-16), 1915-1926. <https://doi.org/10.1016/j.quascirev.2007.06.007>

**Copyright**

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

**Take-down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

# Building and testing age models for radiocarbon dates in Lateglacial and Early Holocene sediments

S.P.E. Blockley<sup>a,\*</sup>, M. Blaauw<sup>b,1</sup>, C. Bronk Ramsey<sup>a</sup>, J. van der Plicht<sup>c,d</sup>

<sup>a</sup>*Research Laboratory for Archaeology, University of Oxford, Oxford OX1 3QY, UK*

<sup>b</sup>*Center for Mathematical Research CIMAT, Guanajuato, Mexico*

<sup>c</sup>*Centre for Isotope Research, Groningen University, Groningen, The Netherlands*

<sup>d</sup>*Faculty of Archaeology, Leiden University, Leiden, The Netherlands*

Received 20 April 2007; accepted 19 June 2007

## Abstract

The growing importance of understanding past abrupt climate variability at a regional and global scale has led to the realisation that independent chronologies of past environmental change need to be compared between various archives. This has in turn led to attempts at significant improvements in the required precision at which records can be dated. Radiocarbon dating is still the most prominent method for dating organic material from terrestrial and marine archives, and as such many of the recent developments in improving precision have been aimed at this technique. These include: (1) selection of the most suitable datable fractions within a record, (2) the development of better calibration curves, and (3) more precise age modelling techniques. While much attention has been focussed on the first two items, testing the possibilities of the relatively new age modelling approaches has not received much attention. Here, we test the potential for methods designed to significantly improve precision in radiocarbon-based age models, wiggle match dating and various forms of Bayesian analyses. We demonstrate that while all of the methods can perform very well, in some scenarios, caution must be taken when applying them. It appears that an integrated approach is required in real life dating situations where more than one model is applied, with strict error calculation, and with the integration of radiocarbon data with sedimentological analyses of site formation processes.

© 2007 Elsevier Ltd. All rights reserved.

## 1. Introduction

There is significant scientific interest in understanding the nature and timing of regional and global responses to abrupt climate change during the transition from the last glacial to the Holocene, due to the numerous abrupt climatic oscillations that are now known to occur during this period as indicated by the Greenland ice core records (e.g. Alley et al., 1993). These are seen as essential natural archives for understanding climate forcing mechanisms and the underlying Earth system response (Walker et al., 2003). In order to understand these processes, it is necessary to compare different archives at sufficiently high temporal

resolution to examine the timing of response between them. The abrupt nature of these climatic events, however, and the possibilities of asynchronous responses in different environments, along with regional climate gradients, requires significant improvements in the precision at which many environmental records can be dated (Lowe et al., 2001). In particular, there have been recent attempts to build age models for radiocarbon-dated sediment sequences with levels of precision normally associated with annual chronologies, such as the Greenland ice cores and dendrochronology. These include visual wiggle match dating of terrestrial sediment records, regression equations applied to dated sequences after calibration, and the application of various Bayesian statistical models to the calibration of dates (e.g. Blaauw et al., 2004; Blockley et al., 2004). All of these methods are, in fact, models of the underlying true sedimentation rate, which are based on various assumptions about the sedimentary archive, the

\*Corresponding author. Tel.: +44 1865285202.

E-mail address: [simon.blockley@rlaha.ox.ac.uk](mailto:simon.blockley@rlaha.ox.ac.uk) (S.P.E. Blockley).

<sup>1</sup>Present address: Department of Archaeology and Palaeoecology, Queen's University Belfast, Belfast, UK.

radiocarbon dates, and the radiocarbon calibration curve itself. Some of these models are *ad hoc*, such as the visual wiggle match adopted by some authors, while others are based on classical or Bayesian statistical frameworks. Such models raise questions of genuine importance for many researchers involved in understanding the nature and timing of climate change in the Lateglacial and Early Holocene, regarding the reliability of the assumptions made in developing these models and their ability to accurately predict the depositional history of a given site. These questions can be examined to some extent by simulating different dating scenarios (Telford et al., 2004). Here, we compare the performance of three of the most common approaches, in four different scenarios that we believe to be representative of the range of sedimentary regimes and dating difficulties. One assumption that cannot be tested in this way, however, is whether the available radiocarbon calibration curve is a completely accurate representation of the terrestrial/atmospheric radiocarbon activity of the whole period through the Lateglacial and into the Holocene. The presently recommended calibration curve is IntCal04 (Reimer et al., 2004), and is based on dendrochronologically dated wood back to 12 460 cal BP. Thus, for this time-range it represents the natural atmospheric/terrestrial  $^{14}\text{C}$  content. Beyond the dendrochronological part, in the earliest part of the Lateglacial record, the IntCal04 curve is marine-based.

## 2. Uncertainty in radiocarbon dating

The chief sources of uncertainty in radiocarbon dating can be broken down into three main areas, which have been covered extensively in the literature (see e.g. Lowe et al., 2001): (1) the reliability of individual radiocarbon dates, which is often difficult to assess in many contexts, owing to a number of issues outlined below; (2) the reliability of the available calibration curve for a given period; and (3) the shape of the calibration curve, which leads to increased errors on calibrated radiocarbon dates and non-normally distributed probability functions. For any methodology to be successful in improving the precision of calibrated radiocarbon chronologies, while retaining accuracy, all three of these problems must be addressed.

### 2.1. Radiocarbon quality assurance

In marine and terrestrial contexts, the potential for individual radiocarbon dates to yield incorrect results have been reviewed extensively in the literature (e.g., Shore et al., 1995; Wohlfarth et al., 1998). Problems include mislabelling during subsampling, non-removed *in situ* or laboratory contamination, marine and lake reservoir offsets, humic acid percolation down a sequence, reworking of macro-fossils within sediments, and detrital carbonate contamination. Given these problems, it is unsurprising that in most cases there are dates which are deemed to be outliers in

most studies where intensive radiocarbon dating has occurred (see e.g. Lowe and Walker, 2000). These are not necessarily a problem if a small number of outliers can be successfully detected, either statistically or through some screening approach. Recent work has involved a combination of a radiocarbon inventory approach (Walker et al., 2003), with some method of outlier detection. Nevertheless, there is potential for misleading ages to significantly skew age models from calibrated radiocarbon dates if outliers are not detected. Intercomparison is a major part of quality assurance, where the same samples are measured by different laboratories. Issues such as conventional vs. AMS dating, quantification of variations and pretreatment can be investigated. For details, we refer to a complete issue of Radiocarbon dedicated to such issues (Scott, 2003). A summary of the findings has been published in JQS (Boaretto et al., 2002).

Several studies have now shown that the reliability of dating these various fractions of sediment is site specific, and that where possible assessing the true ‘radiocarbon inventory’ of a site is the only real way of establishing the most robust data-set possible for any age modelling. The models that are examined below are only as effective as the data presented, and it is only with a radiocarbon inventory that this can be evaluated. For example, dating at the site of Llanilid in South Wales revealed that while there was considerable disagreement between dates from the same horizons, the main discrepancy in dates at the site was between humin ages and all other fractions, and in this case a coherent site chronology was developed from a combination of humic and macro-fossil ages (Walker et al., 2003), which allowed the development of a Bayesian age model (Blockley et al., 2004).

### 2.2. Radiocarbon calibration

The reliability of any radiocarbon-dated age model rests considerably on the reliability of the available radiocarbon calibration curve. Without calibration, radiocarbon chronologies cannot be reliably related to non-radiocarbon-based data, and it is highly inadvisable to examine changes over time on the significantly non-linear radiocarbon timescale. Recent work undertaken in the development of the new IntCal04 calibration curve suggests that for the period in question IntCal04 is a reasonable model of the underlying true calibration curve, with improved data sets and formal statistical modelling of the underlying calibration curve from the available data (Buck and Blackwell, 2004; Reimer et al., 2004). Nevertheless, for the Lateglacial concerns remain over the effects of changes in marine reservoir offset on the marine-based section of the curve beyond 12 460 cal BP. The calibration curve has a decadal temporal resolution through the tree-ring-based section of the curve, moving to a centennial resolution in the marine-based section, leaving aside reservoir offset concerns. Once radiocarbon dates have been calibrated, however, the resultant ages have large non-normally distributed ranges.

These data do not lend themselves easily to statistical analyses, and traditional regression-based approaches to age modelling are highly unsuited to dealing with non-normally distributed data, especially with large uncertainties on each data point. Increased radiocarbon precision can help, but in some parts of the curve even the highest resolution radiocarbon dates with impossibly low laboratory errors will still result in calibrated errors in the range of hundreds of years; the most illustrative example occurs during the “Hallstatt” plateau (ca 800–400 BC).

### 3. Testing age modelling approaches

Any methodology for modelling sedimentation history based on calibrated radiocarbon dates, be they *ad hoc* or a formal statistical framework, needs to demonstrably improve upon the precision available by simply calibrating radiocarbon dates on their own, while at the same time retaining accuracy, and reliably dealing with outliers. If we assume that the IntCal04 calibration curve is indeed reliable for the period in question, then it is possible to test any available approach by generating artificially simulated sedimentation regimes, and using the calibration curve to simulate radiocarbon dates from them. This method was used by Telford et al. (2004), to successfully test the use of regression equations to model sedimentation rates, which demonstrated that, in most cases, the number of radiocarbon dates used in age models of this type was far too small. Here, we adopt this approach to test other methodologies that are becoming prominent in the literature: wiggle match dating and Bayesian modelling.

#### 3.1. The models

The models tested, outlined in more detail below, have been broken down into the following two groups. Firstly, we consider uniform depositional models that, within predefined zones, assume a uniform sedimentation rate and, therefore, attempt to match the structure of the radiocarbon dates, when plotted against depth, to the shape of the radiocarbon calibration curve. This process is generally termed ‘wiggle match dating’. Within this group there are two approaches tested: the visual wiggle match that attempts this process in a relatively *ad hoc* manner, which makes error quantification difficult; followed by two forms of Bayesian statistical matching of the data to the calibration curve: Bayesian wiggle matching in the Bpeat software, and in one comparability test, the similar model in the OxCal software (*U\_Sequence*), both termed here Bayesian wiggle match dating (BWMD). We, secondly, consider the potential for modelling sequences using less constrained Bayesian statistical methods (termed here Bayesian sequence modelling), where the depositional rate is allowed to move away from uniformity within the defined stratigraphical zones (*Sequence* and *P\_Sequence*). The aim is to outline a practical methodology for the community when applying these models.

#### 3.2. The scenarios

Four different sedimentary scenarios were devised from simple to complex. A series of calendar ages for these scenarios were simulated with spacing of between 50 and 200 years. This relatively close chronological spacing was chosen in the light of recent studies highlighting the need for a large corpus of radiocarbon dates, in order to reliably build age models for a site (Walker et al., 2003; Telford et al., 2004). In actuality, most of the techniques would have worked perfectly well with fewer dates, as long as there were enough dates to reliably define inflection points in the sedimentation rate. As, however, these are rarely known in detail before a dating study begins, it is increasingly the practice within the community to use large numbers of radiocarbon dates to search for changes in sedimentation rate. Once calendar ages are generated they can be used to simulate radiocarbon dates. This can be performed in a number of ways, either manually or through modelling the relationship between calendar ages and the calibration curve. We have used the simulation function in the OxCal package, which generates a radiocarbon date (with randomly sampled but realistic errors), given the calendar age expected and the laboratory error term that, in this case, we have selected as 50 years, based on our observations of the levels of precision regularly achievable by many laboratories when dating samples in the Lateglacial and Early Holocene. Once generated the radiocarbon ages were given blind to other members of the team, to attempt the reconstruction of the calendar sedimentation rates using one of the methods described above. The operators of the software were only given information that could have been gleaned from observing cores, such as the position of hypothesised lithostratigraphic boundaries that may indicate changes in sedimentation rate. From these data, they were asked to reconstruct sedimentation rates, within the model errors of the technique being employed, for the following four scenarios:

1. Here, we devised a simple simulation of a site covering the period 13 000–9000 calendar years BP. The simulation site had a uniform linear sedimentation for the whole period, and in the simulation the site was dated every 100 years. Each of these calendar ages can then be converted to radiocarbon dates, as described above with, in this case, an assumed 50 year error on the precision of each simulated date. This is the default case for the wiggle match approach, as it meets all of the a priori assumptions of the technique. The sedimentation rate is simple to reconstruct and there are numerous simulated radiocarbon dates, >95% of which correctly predict the simulated age–depth relationship when calibrated at 95% confidence, prior to any modelling. We expected all of the methods to perform well in this scenario in terms of accuracy, with wiggle matching methods giving the highest precision.

2. This simulation introduces levels of complexity that are often found in sequences that cover long periods of time, with several different sedimentation rates. We have mimicked a typical tripartite sedimentation rate that has been reported from several Lateglacial sequences (e.g. Walker et al., 2003), and added another change of sedimentation in the later part of the core. Although this may seem quite complex, it reflects evidence from the literature, and all four of the sedimentation units were linear; this was seen as a good test of all the methods in difficult circumstances. As with the case above, there were numerous simulated radiocarbon dates and >95% of them correctly predicted the simulated age–depth relationships when calibrated at 95% confidence, prior to any modelling.
3. This simulation was based on the idea that sedimentation rates could be generally linear, but with deviations from linearity during transitions between sedimentary phases, and such depositional regimes have been recorded in independently dated varved deposits, used in other tests of radiocarbon-based age modelling (e.g. Telford et al., 2004). We were interested to observe how well each technique coped with these transitions, and, particularly with wiggle match dating, how well the software was able to detect these deviations, approximating them with a number of linear functions.
4. This simulation was identical to simulation 3, but we introduced ~10% outliers, as we wanted to test how well the models coped with the detection and rejection of outliers, and also how badly the resulting age models would have been affected if the outliers were not detected.

Using these simulations, we assess the reliability of the different approaches to age modelling that are described above. We assess them in terms of their precision and accuracy; with good model performance being deemed as an increase in precision, beyond the error associated with simply calibrating the radiocarbon dates alone, while retaining accuracy in predicting the true simulated sediment age, within the model errors of the technique employed. Thus, we discuss the uncertainty calculated by each technique, which is a measure of how well the technique improves upon the precision of the calibrated date; and the accuracy, which is the difference between the simulated true age and the age range calculated by a particular model. All of the results of the simulations below are reported in Appendix A, where not highlighted in the text.

#### 4. Simulation tests

##### 4.1. Visual wiggle match dating

The term wiggle match dating is rather loose in its application and can lead to confusion. Many wiggle matches are based on the radiocarbon dating of independent, floating, but annually resolved records, such as tree

rings. These are the most robust forms of wiggle match, and software to perform these analyses has been available for a number of years. There are, however, other forms of wiggle matching, where there is no independent chronology, and where the spacing between the dates has to be inferred from some assumption of the sedimentation rate. The best match between the shape of the calibration curve and the dates in the sequence has to be found by gradual alteration of the, initially unknown, sedimentation rate. For a number of years this approach has been used successfully to wiggle match across short periods of time (a few hundred years), to derive dates for events such as tephra depositions (e.g. Plunkett et al., 2004). Wiggle matches have also been attempted across much longer stratigraphical units, such the Lateglacial and Early Holocene, often without error calculations being attempted (for a discussion of the issues, see Davies et al., 2004). It is specifically these types of wiggle match that are most pertinent to this study.

We attempted first to ‘visually wiggle match’ the simulated dates to the calibration curve, by eye, and use them to reconstruct the simulated sedimentation rates. As the dates were themselves constructed from the calibration curve, we believed that if visual wiggle match dating is performing as well as it is generally believed, then at least in the first two scenarios it should be more than capable of reconstructing the sedimentation rates. One problem for the method, however, is that it is somewhat subjective and is often performed on the assumption that the least number of segmentations that provides a reasonable visual fit is the best chronology. We anticipated, therefore, that in scenarios three and four, where there are some complexities in the transitions between different linear sedimentation rates, the visual wiggle match would break down. Several visual attempts were made, all giving different answers, yielded by different researchers. The best results, however, were achieved using a simple piece of software developed by Blaauw (see Blaauw et al., 2003) for performing visual wiggle matching in a spreadsheet package. This performs the analyses by assuming a linear accumulation rate for a core, choosing ‘reasonable’ initial values for the two parameters involved (accumulation rate and start calendar year), and translating the dated depths into calendar ages through these parameters. Then the  $^{14}\text{C}$  dates are plotted together with the calibration curve, to check how well the sequence fits. After this the two parameters are adapted manually. The accumulation rate is used to compress or expand the sequence, and the start calendar year to move the sequence to the left or to the right, until a reasonable wiggle match of the dates with the calibration curve is found. When entire sequences cannot be matched with satisfaction, subdivisions are inferred manually and matched individually.

##### 4.2. Performance

*Scenario 1:* Here this method performed extremely well, reconstructing the linear sedimentation rate exactly (see



Appendix A). As the radiocarbon dates used were generated by using the calibration curve, the absolute precision is a little false, and in reality the errors on the calibration curve itself would have to be included in the error calculation, which is not always attempted in visual wiggle match dating. Nevertheless, in a very simple scenario, with all assumptions being met, visual wiggle match dating does offer the precision and accuracy that has been suggested it is capable of.

**Scenario 2:** The first scenario is unlikely to represent the depositional history of many records that cross from the Lateglacial through to the Holocene, and we believed that the second scenario would be a more interesting test of the potential for visual wiggle match dating to model a series of linear sedimentation rates. While the method produced linear sedimentation rates for much of the sequence that were within  $\sim 20$  years of the simulated sedimentation rates (Fig. 1), the method failed where there were sharp changes in sedimentation rate, yielding errors in excess of 100–150 years.

**Scenarios 3 and 4:** These were devised to test the ability of all models to examine sequences that have non-linear changes in sedimentation rate at periods of environmental forcing, and, in the case of scenario 4, containing a percentage of outliers. It was hoped that the wiggle match methods would be able to approximate these with a series of small linear models, and that they would be able to still successfully model the stable linear parts of the simulation. While the visual wiggle match performed well in the stable parts of the sequence, with inaccuracies (difference between modelled and true age) running at  $\sim 30$  years for scenario 3 and  $\sim 50$  years for scenario 4, the visual wiggle match was unable in either case to suitably model the complex depositional changes, showing inaccuracies in the order

of  $\sim 150$  years for scenario 3 and 150–400 years for scenario 4 (Fig. 2).

One definition of success in these tests is the ability to accurately reconstruct the simulated sedimentation rates, within errors. As visual wiggle match dating errors are not always computed, in one sense, the visual wiggle match, as applied here, failed all but the first test, which is the idealised case, and for long sequences is very unlikely to be reliable. From our experience, no real core covering such long time periods has ever been as easy to wiggle match as this simple scenario. For example, one of us (Blaauw) has attempted wiggle matches (both visual and modelled) on ca 40–50 cores, and only two allowed ‘reasonable’ fits to the curve without subdivisions into subsets (e.g. at the site of Borchert: van der Plicht et al. (2004), but these wiggle matches are actually not very successful, as several dates do not match very nicely). Interestingly, a core with some 10 dates matched well without needing sections; however, additional dating (40 dates in total) required division into more levels. One disturbing fact was that during the inflection periods, between different linear sedimentation rates, or during periods of fluctuating sedimentation rate, the visual wiggle match performed poorly, being falsely precise, as no errors are estimated, in this case depths are translated into mere point estimates of calendar years. Inaccuracies of hundreds of years results, while returning models that would be deemed acceptable by the standards reported in the literature. This leads us to suggest that those indulging in visual wiggle matching take significant precautions, and assume that, unless they have strong evidence that they are in the idealised case, that there are indeed errors that run constantly between 20 and 50 years, and often be the order of hundreds of years.

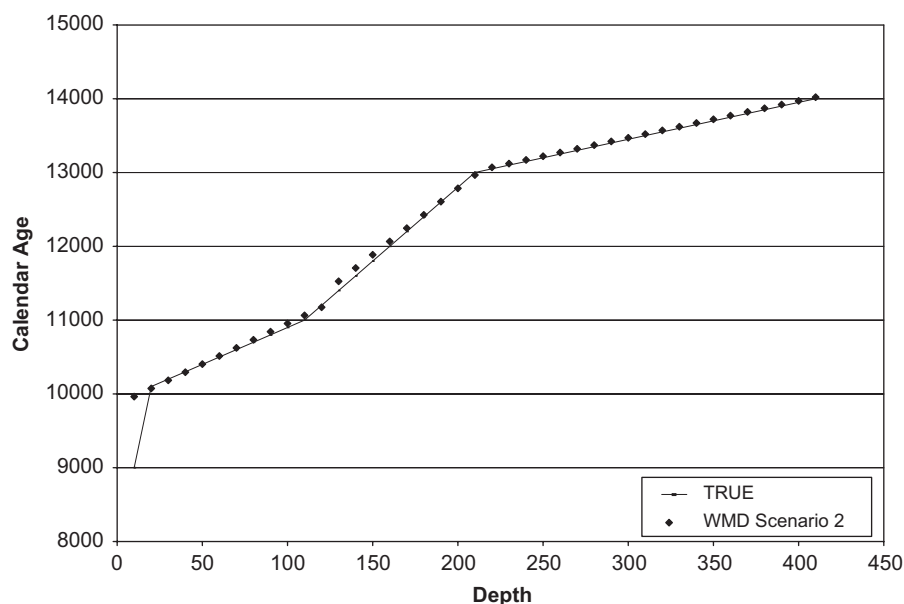


Fig. 1. Visual wiggle match dating results and the expected ('true') results, plotted against depth, for simulation 2. Although the answer is generally correct, the lack of calculated errors means that there is failure in many areas of the reconstruction, even in such a simple test, on absolutely ideal data.

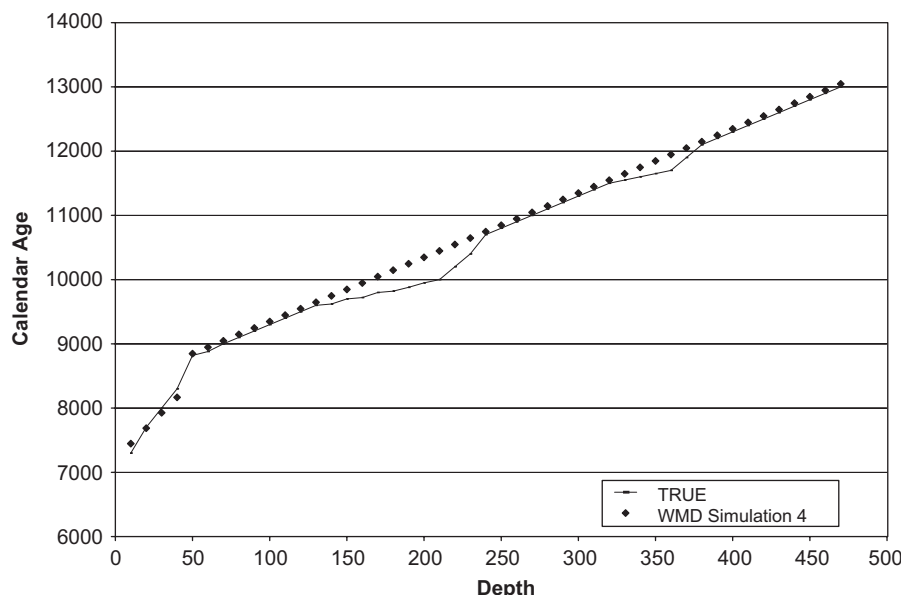


Fig. 2. Visual wiggle match dating and the expected results, plotted against depth, for simulation 4. Here, the method has failed for most of the reconstruction, despite returning a match that would normally be accepted as reasonable.

It should be remembered that in many cases visual or *ad hoc* wiggle matches are performed over much shorter time periods, covering stable environmental conditions with some attempt to compute uncertainties (e.g. Pilcher et al., 1995). It is, therefore, less likely that such wiggle matches will be as open to significant unobserved methodological uncertainties. Nevertheless, the main aim of this paper is to examine the potential to improve chronological reliability over Lateglacial and Holocene timescales, and it is in these cases where there is apparent need for better tools than the visual wiggle match. There are more formal statistical tools for applying and testing wiggle matches, and other age models, based on sedimentary assumptions. These we discuss in turn using several models in two readily available software packages. Firstly, we examine the method-specific Bpeat software, designed to perform BWMD in steadily accumulating environments. We follow by examining the range of assumptions that can be applied to constructing age models in the OxCal software.

#### 4.3. Bayesian wiggle match dating (BWMD)

This approach attempts to resolve the subjectivity of wiggle match dating by statistically identifying the most probable match for a given series of dates, dividing into subsections when necessary, using Markov Chain Monte Carlo simulations. We have mainly relied on the Bpeat software and both the technique and software have been described in detail in Blaauw and Christen (2005). To summarise, this software was designed to create statistical wiggle matches from dates in peat cores, with their specific peculiarities, such as the reasonable assumption of linear accumulation. The software assumes that: (i) at any time, hiatuses or changes in accumulation rate can occur,

necessitating division of  $^{14}\text{C}$ -dated sequences into sections; (ii) within sections accumulation rate is linear, with accumulation rates of 10–20 yr/cm being the most likely, and other positive rates being possible, but less likely; (iii) between sections hiatuses can occur (in bogs for example, owing to fires or bog bursts, in lakes these could be caused by low lake levels), with short-lasting hiatuses being most likely; (iv)  $^{14}\text{C}$  dates have a prior probability of being outlying; and (v)  $^{14}\text{C}$  dates can suffer from a reservoir effect. Through millions of MCMC iterations, the prior information is combined with the dates, resulting in estimates of calendar ages for every depth of a core. The millions of iterations of the included parameters provide an estimate of their posterior distributions, and can be used to provide confidence levels (error bars) for derived calendar ages. The proposed age-model results from the iteration that gave the best fit of the dates to the calibration curve (the *Maximum A priori*; Blaauw and Christen, 2005). Although Bpeat performs well in many of the tests below, it is possible that performance could be improved if the software parameters were altered for other conditions, although in the end linear sequences only can be modelled in this way. In addition, we also ran the comparable Bayesian function in the new version of the OxCal software (*U\_Sequence*), which gave almost identical results on the first model tested, and was subsequently dropped from further analyses, to avoid unnecessary duplication.

As well as being a more formal statistical approach, BWMD has the advantage over visual wiggle matching that errors can be calculated, using the posterior distributions of all parameters involved, and outliers identified, within a Bayesian framework. Outliers are assessed using a prior probability of 5% of dates being incorrect; if a date needs a large ( $>2\text{sd}$ ) shift in its  $^{14}\text{C}$  age to fit the model and

the other data, it is identified as an outlier, and is given less weight in the inference process (see also Christen, 1994). It was hoped that in all scenarios this approach would be able to approximate the simulated sedimentation rates, within errors, using a series of subdivisions. This method, as with all the formal statistical approaches that follow, allows the calculation of model uncertainties, which are always expressed as 95% HPDF, or more formally, as these are Bayesian models, 95% highest probability density functions (HPDF).

#### 4.4. Performance

*Scenario 1:* BWMD performed extremely well in this test, using either method. As with visual wiggle matching, it was able to accurately reconstruct the simple linear sedimentation rate, and, in addition, calculate errors that reflect the uncertainty in the construction of the calibration curve. In ideal conditions, this method would be recommended by us as the most appropriate calibration approach. For the remaining scenarios, only Bpeat was used as both models are mathematically comparable.

*Scenario 2:* BWMD using Bpeat performed well in this test (Fig. 3); it always reconstructed the multiple linear sections within model errors that were in the range of 3–135 years, in most phases of the reconstruction. Only at the top of the sequence did Bpeat fail to reproduce the sedimentation history within errors. This is because Bpeat was designed such that if it encounters a strangely lying date, it can either choose to (1) divide there, and make another subsection [as should have been the case as this was the ‘true’ event]; or (2) assume that the date was an outlier, and not add much importance to this date. In this case, this is what Bpeat chose to do, and a visual wiggle

match would have probably done the same, as there was not much evidence given to the operators suggesting the requirement for another subset. Extra subsets imply additional parameters to estimate and fewer degrees of freedom, Occam’s razor would therefore suggest that the date was an outlier. Extra dates would be required in this case, to show that the date was not outlying, and would instead suggest adding another section. Although all models have particular problems with dates such as this, especially at the beginning or end of sequences, it appears to have particularly affected the wiggle match methods. The accuracy of predicting the real calendar ages overall, however, is 91%, making this method the joint top performer in this test, along with another Bayesian method the *P\_Sequence*, outlined below.

*Scenarios 3 and 4:* The method performed reasonably well in both of these scenarios, and in general reconstructed the age–depth relationships, within the calculated errors. In simulation 3, the error estimated by the Bayesian matching was in the order of ~60 years for most of the core (see Appendix A). The difference between the predicted wiggle match age and the ‘true’ age was in fact under around 50 years for most of the core and rose to ~200–400 years at the deviations from linear sedimentation. In simulation 4, the method reconstructed the age–depth relationships with an average error range of 73 years, this value again being skewed by the ages at the deviations from linear sedimentation. In both cases, the method missed one or more of the deviations from a linear sedimentation rate (Fig. 4) and modelled a statistically acceptable but incorrect linear sequence though a non-linear sedimentation. For section 4, the inbuilt outlier detection system appeared to successfully negotiate the inbuilt false dates.

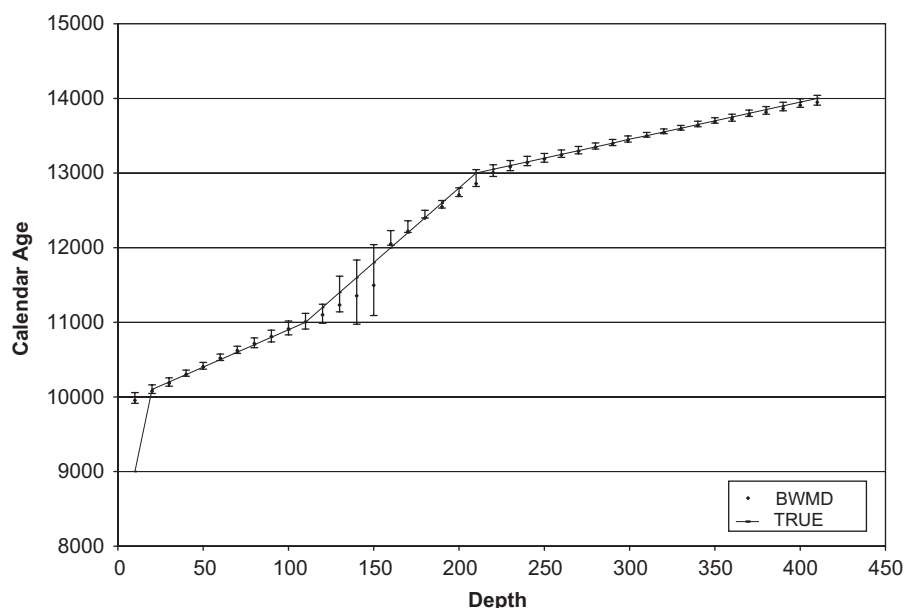


Fig. 3. Bayesian wiggle match (BWMD) for simulation 2. Here, the method is reasonably accurate and errors are small. This along with the *P\_Sequence* was one of the best performers for this simulation.



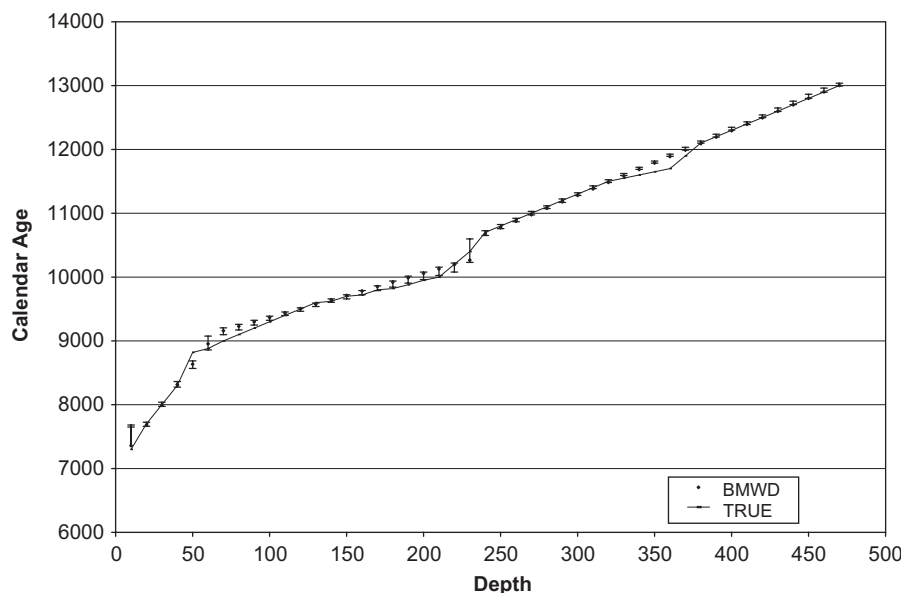


Fig. 4. Bayesian wiggle match (BWMD) results for simulation 4. Here, the method is correct for much of this sequence, but difficulties have been encountered with the non-linear sections.

#### 4.5. Bayesian sequence modelling and the OxCal software

Although the wiggle match dating method described above is Bayesian, sharing many of the basic underlying mathematics and assumptions with other Bayesian age–depth modelling software that is on general release, we have separated it from the general discussion of Bayesian sequence modelling, arbitrarily, as it is designed as an improvement on wiggle match dating in environments where linear sedimentation rate assumptions and relatively low sedimentation rates are likely. For these, it is ideally suited as indicated by the performance in scenarios 1 and 2. This requirement for linearity of sedimentation rate is not a feature of the remainder of the Bayesian approaches we test. Here, what we aim to test are Bayesian models in general that use the uniform prior (see below) and the applicability to a range of scenarios of the more flexible Bayesian models that allow non-linear sedimentation rates across a sequence.

Until recently most Bayesian methods developed to model age–depth relationships for radiocarbon dates have been built around the uniform prior, sometimes termed the *Sequence* algorithm (see Steier and Rom, 2000) in sedimentary sequences. In this, a series of assumptions are modelled during the calibration process. Although there are differences between different packages in some details, such as the methodology adopted for outlier detection, all have the majority of assumptions in common. These methods have been discussed widely (Buck et al., 1991, 1992; Bronk Ramsey, 1999, 2000, 2001; Steier and Rom, 2000; Blockley et al., 2004) and need no repetition here, but it is important to summarise the key assumptions included in the uniform prior model for calibrating dates in sequence. These are, firstly, that by using the structure of

the calibration curve itself areas of more or less likely calendar age can be developed within the overall range of a date and, thus, a probability density function for a calibrated date can be derived. Secondly, that the law of superposition (age should increase with depth) can be imposed upon the probability densities (priors) of a group of dates in a sequence during calibration to restrict the most likely range of the resulting calibrated ages (posteriors) of all the dates. This is based on each individual date and the positions of all of the others in stratigraphical order and time. Finally, that it is reasonable to assume that where dates are spaced closely enough within a sequence for this to have an effect upon the range of the calibrated ages then an assumption of linear sedimentation between each date is reasonable. To give a simple example, if five radiocarbon ages have a series of calendar ages  $a_1, \dots, a_5$ , with the stratigraphical relationship, oldest first,  $a_1, a_2, a_3, a_4, a_5$ , assuming stratigraphy and succession (law of superposition), the prior assumption ( $P$ ) of their chronological relationship would be

$$P = a_1 > a_2 > a_3 > a_4 > a_5.$$

The applicability of the uniform prior to sedimentary archives has been discussed elsewhere (Bronk Ramsey, 2000; Steier and Rom, 2000; Blackwell and Buck, 2003) but while it is possible that the uniform prior can underestimate the full range of possible ages of calibrated dates in a sequence, due to the possibility of non-linear sedimentation rates between individual calibrated dates, it has been shown that with a reasonable number of dates and appropriate nesting of sequences within boundaries, that the method can reliably reconstruct sedimentation rates within model uncertainty that are significantly smaller than if the dates were calibrated on their own.

Nevertheless, the use of the uniform prior has never been tested on varying sedimentation rates in an exhaustive a sense as the test described here and given the increased use of the method (e.g. Walker et al., 2003; Blockley et al., 2004) it was deemed useful to apply this methodology.

In addition to the traditional *Sequence* model used widely in different software (e.g. BCAL, OxCal), additional statistical models are being developed to improve the performance of the Bayesian methodology in different circumstances. These are being developed for the new release of the calibration package OxCal and the most innovative is tested here, the *P\_Sequence* (Bronk Ramsey, in press). OxCal4 now offers potential improvements in both the modelling of age–depth relationships using posterior distributions after Bayesian calibration with the uniform prior. As such we wished to test if these newly developing methods would improve upon the simple application of the *Sequence* algorithm. We believe such age modelling may necessary, as it is often the case that once dates have been calibrated using a Bayesian approach inappropriate secondary age modelling takes place, where regression equations are applied to the age–depth data, with some point estimate of the calibrated age being used. Given the non-normal and often bi-modal probability distributions of calibrated ages, even after Bayesian analyses, this is far from ideal. The *P\_Sequence* model is seen as a general improvement on using the *Sequence* algorithm alone and was applied to all of the models. The reason for the perceived improvement is that this method also utilises the depth information for the radiocarbon dates as well as the laws of stratigraphy and succession in constraining the posterior age ranges of the dates. In utilising the depth relationships between dates while not imposing a uniform linear sedimentation rate for the

sequence, this approach is a halfway house between the general *Sequence* algorithm and BWMD.

#### 4.6. Performance

*Scenario 1:* As expected both models performed well, accurately predicting the simulated age–depth relationship and improving upon the error ranges achievable with calibration of the dates alone. Performance also followed as expected the level of increased forcing on the model, with the standard sequence model being the least precise, followed by the *P\_Sequence*.

*Scenario 2:* Both of the models used performed well, returning 100% accuracy and model error terms consistently lower than those associated with just calibrating dates on their own. The *P\_Sequence* was the best performer of the two (Fig. 5) on this test with average model error ranges of 69 years compared to an average model error of range of 123 years for the *Sequence* model.

*Scenarios 3 and 4:* The two methods performed reasonably well in these harder tests (see Appendix A) with the *Sequence* model performing best in terms of accuracy while the *P\_Sequence* was more precise. For simulation 3, the general sequence model reconstructed the age–depth relationship with an accuracy of 98%, however, precision was lowered and the average 95% (HPDF) error range was 228 years. The *P\_Sequence* achieved 99% accuracy with an average 95% HPDF error range of 145 years. As with the Bayesian wiggle match approach, however, the method failed to reconstruct one of the major deviations in sedimentation rate. For simulation four, which is the same data as simulation 3 but with 10% outliers, the *Sequence* model negotiated these outliers well retaining an accuracy of >99%, accurately reconstructing the age–depth

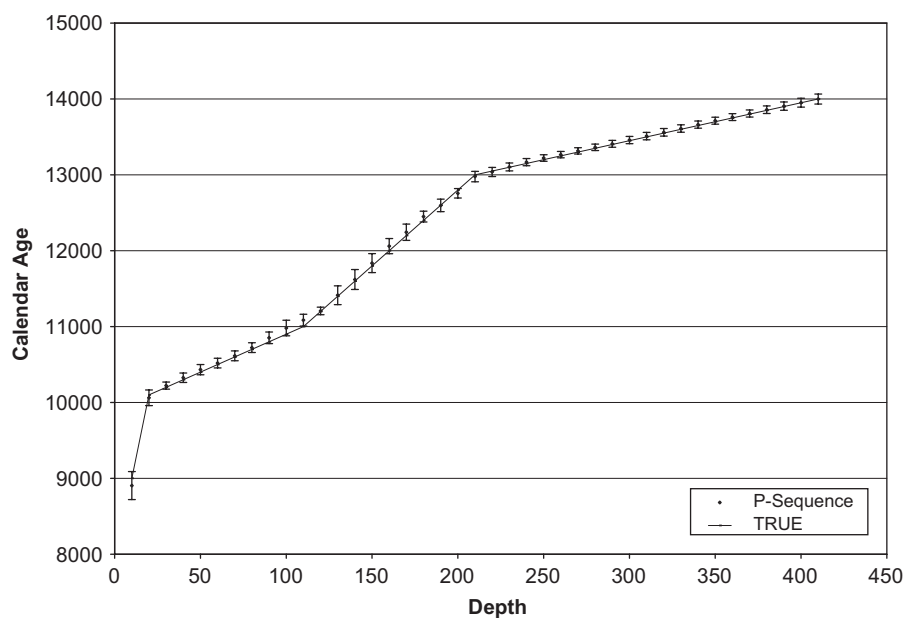


Fig. 5. OxCal *P\_Sequence* results for simulation 2, here the simulation is 100% correct and errors are a clear improvement on just using calibrated ages alone.

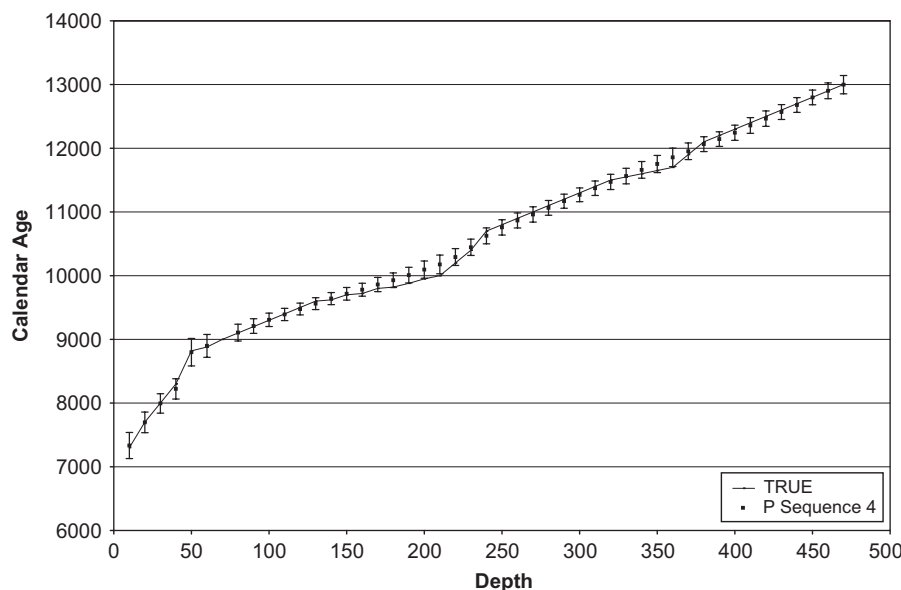


Fig. 6. OxCal *P\_Sequence* results for simulation 4, here the method is 95% accurate and has only failed to negotiate one deviation from a linear sedimentation rate.

relationship with an average 95% HPDF error range of 224 years Appendix A. Statistically the *P\_Sequence* performed equally as well in this test with an accuracy of 96% and an average 95% HPDF error range of 255 years; however, crucially it failed, along with most other models, to fully model one of the deviations in sedimentation rate at around 360 cm in depth (Fig. 6).

## 5. Discussion and suggestions for approaches to age modelling

When attempting to draw conclusions from these tests that are useful for the wider community it must be remembered that all of the dates used were simulated from the calibration curve. As such, they are idealised and the normal possibilities of time averaging and reworking are generally not present in these cases, the introduction of 10% outliers in simulation four notwithstanding. This means that caution must always be applied when selecting a particular model approach, and clearly these methods will only work with a sufficiently large number of dates that are themselves correct.

From the model results three points are apparent. Firstly, that the two least constrained Bayesian models (*Sequence* and *P\_Sequence*) can give generally reliable improvements in precision, in a range of conditions, providing the dates are mostly correct. Secondly, where the underlying assumptions are met, with a single linear sedimentation rate, the wiggle match approach, using either an *ad hoc* or Bayesian framework, can generate very high precision while retaining accuracy, and it is only the error on calibration and the reliability of the calibration curve itself that limits the method in these ideal circumstances. Finally, that the Bayesian wiggle match method

can produce reasonable approximations within errors in the second scenario of segmented linear sequences.

From the above observations, it is possible to suggest that users of these methodologies would take value in attempting at least two models (preferably using at least one of the less constrained Bayesian models first), as no one method is ideal for all circumstances in terms of maximising precision whilst retaining accuracy, although the *P\_Sequence* is the overall best performer across all models. While trying more than one model seems logical it is the general statistical trend to choose the model first, and there is a danger when applying several models that the research would choose one that gave a preferred result for pre-conceived notions of expected results; for example, the ‘suck in and smear out syndrome’ (Baillie, 1991). We would suggest, therefore, that the main justification for choosing a particular age model should be based wholly on internally demonstrable reliability.

The modelling should only be one part of the dating strategy, if the highest possible resolution for a given site is aimed at. Implementing these models into the dating strategy should begin with thorough analyses of the sedimentology, lithostratigraphy and micromorphology of the site in question. As our experiments demonstrate, there is little point in wiggle matching a site that is likely to have had significant deviations from linear sedimentation rate and sporadic rapid inwash of sediment; however, in the conditions Bpeat was designed for (Scenario’s 1 and 2) this is an ideal approach. Once a site has been understood in these terms the radiocarbon inventory of the site must be examined. As has been demonstrated (Walker et al., 2003; Blockley et al., 2004), a combination of a radiocarbon inventory and the least constrained of the Bayesian models (*Sequence*) can be used to find the most robust data set of

radiocarbon dates for modelling. Beyond this, once a coherent data set has been developed, as a guide to the general user they should start with the *P\_Sequence*. This should let users find the shape of the age–depth data in the sequence. From this, in sediments that are suited to wiggle match dating (peats particularly), assuming the *P\_Sequence* does not show non-linearity, as can be seen from tests 3 and 4, they should go on to try a wiggle match, using Bpeat. If there is a more complex age–depth profile revealed by the *P\_Sequence*, with areas that could not be modelled with one or more linear wiggle matches then they should stop there, as the wiggle match assumption has broken down. Where users do find a wiggle match possible we would recommend Bpeat as there is an inbuilt ability to calculate errors and calculate convergence, or the similar *U\_Sequence* model in OxCal. Such ability is vital in producing reliable chronologies, as is displayed clearly where the visual wiggle match without errors is incorrect the majority of the time.

## 6. Conclusion

These series of tests have shown that, in theory at least, in ideal scenarios, the high precision proposed by proponents of wiggle match dating and Bayesian modelling is achievable, and that even in quite difficult simulated scenarios at least centennial chronological resolution is achievable. However, these studies highlight the caution that must be used in the production and interpretation of these models. Key points are:

1. All age models have errors, even the most precise and effective wiggle match should at least include the errors on the calibration curve from which it is derived, as in reality the available calibration curve is itself only a model of the underlying truth of radiocarbon production in the atmosphere (Buck and Blackwell, 2004).
2. Once sedimentation rates move from the very simple case, outlined in simulation 1, there will be errors up to the centennial scale on any wiggle match, at least at the boundary between different linear sedimentation rates, these errors exist whether they are calculated or not.
3. In very complex circumstances, the less precise but more flexible Bayesian sequence models provide the most robust chronologies, whereas in more simple sedimentation regimes the Bayesian forms of wiggle match are the most robust performers.
4. When the different age model approaches are used together problems can be properly addressed. The initial use of the flexible Bayesian sequence models, to examine the likely shape of the sedimentation at a given site, can inform the user of the advisability of proceeding to more precise approaches. We would then recommend the use of the Bayesian wiggle match software tested here, as there is the inbuilt function of error calculation. If these errors become large then it is unlikely that a site's sedimentation history is suitable for a wiggle match approach.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.quascirev.2007.06.007](https://doi.org/10.1016/j.quascirev.2007.06.007).

## References

- Alley, R.B., Meese, D.A., Shuman, C.A., Gow, A.J., Taylor, K.C., Grottes, P.M., White, J.W.C., Ram, M., Waddington, E.D., Mayewski, P.A., Zielinski, G.A., 1993. Abrupt increase in Greenland snow accumulation at the end of the Younger Dryas event. *Nature* 362, 527–529.
- Baillie, M.G.L., 1991. Suck in and smear-two related chronological problems for the 90s. *Journal of Theoretical Archaeology* 2, 12–16.
- Blaauw, M., Christen, J.A., 2005. Radiocarbon peat chronologies and environmental change. *Applied Statistics* 54, 805–816.
- Blaauw, M., Heuvelink, G.B.M., Mauquoy, D., van der Plicht, J., van Geel, B., 2003. A numerical approach to C-14 wiggle-match dating of organic deposits: best fits and confidence intervals. *Quaternary Science Reviews* 22, 1485–1500.
- Blaauw, M., van der Plicht, J., van Geel, B., 2004. Radiocarbon dating of bulk peat samples from raised bogs: nonexistence of a previously reported 'reservoir effect'? *Quaternary Science Reviews* 23, 1537–1542.
- Blackwell, P.G., Buck, C.E., 2003. Space time modelling of Late Glacial re-occupation in Europe. *Antiquity* 77, 232–240.
- Blockley, S., Lowe, J.J., Walker, M.J.C., Asioli, A., Trincardi, F., Coope, G.R., Donahue, R.E., Pollard, A.M., 2004. Bayesian analysis of radiocarbon chronologies: examples from the European Late-glacial. *Journal of Quaternary Science* 19, 159–175.
- Boaretto, E., Bryant, C., Carmi, I., Cook, G., Gulliksen, S., Harkness, D., Heinemeier, J., McClure, J., McGee, E., Naysmith, P., Possnert, G., Scott, M., van der Plicht, H., van Strydonck, M., 2002. Summary findings of the fourth international radiocarbon intercomparison (FIRI) (1998–2001). *Journal of Quaternary Science* 17, 633–637.
- Bronk Ramsey, C., 1999. OxCal. Radiocarbon Calibration and Stratigraphic Analysis Program, Research Laboratory for Archaeology. Oxford University, Oxford <[www.rlaha.ox.ac.uk](http://www.rlaha.ox.ac.uk)>.
- Bronk Ramsey, C., 2000. Comment on the use of Bayesian statistics for C-14 dates of chronologically ordered samples: a critical analysis. *Radiocarbon* 42, 199–202.
- Bronk Ramsey, C., 2001. Development of the radiocarbon calibration program. *Radiocarbon* 43, 355–363.
- Bronk Ramsey, C., in press. Deposition Models for Chronological Records. *Quaternary Science Reviews*.
- Buck, C.E., Blackwell, P.G., 2004. Formal statistical models for estimating radiocarbon calibration curves. *Radiocarbon* 46, 1093–1102.
- Buck, C.E., Kenworthy, J.B., Litton, C.D., Smith, A.F.M., 1991. Combining archaeological and radiocarbon information: a Bayesian approach to calibration. *Antiquity* 65, 808–821.
- Buck, C.E., Litton, C.D., Smith, A.F.M., 1992. Calibration of radiocarbon results pertaining to related archaeological events. *Journal of Archaeological Science* 19, 497–512.
- Christen, J.A., 1994. Summarising a set of radiocarbon determinations: a robust approach. *Applied Statistics* 43, 489–503.
- Davies, S.M., Wohlfarth, B., Wastegård, S., Andersson, M., Blockley, S.P.E., Possnert, G., 2004. Were there two Borrobol Tephra during the early Lateglacial period: implications for tephrochronology? *Quaternary Science Reviews* 23, 581–589.
- Lowe, J.J., Walker, M.J.C., 2000. Radiocarbon dating the last glacial–interglacial transition (14–9 <sup>14</sup>C ka BP) in terrestrial and marine records: the need for new quality assurance protocols. *Radiocarbon* 42, 53–68.
- Lowe, J.J., Hoek, W.Z., INTIMATE group, 2001. Inter-regional correlation of palaeoclimatic records for the last Glacial–Interglacial transition: a protocol for improved precision recommended by the

- INTIMATE project group. *Quaternary Science Reviews* 20, 1175–1187.
- Pilcher, J.R., Hall, V.A., McCormac, F.G., 1995. Dates of Holocene Icelandic volcanic-eruptions from tephra layers in Irish peats. *Holocene* 5, 103–110.
- Plunkett, G.M., Pilcher, J.R., McCormac, F.G., Hall, V.A., 2004. New dates for first millennium BC tephra isochrones in Ireland. *Holocene* 14, 780–786.
- Reimer, P.J., Baillie, M.G.L., Bard, E., Bayliss, A., Beck, J.W., Bertrand, C.J.H., Blackwell, P.G., Buck, C.E., Burr, G.S., Cutler, K.B., Damon, P.E., Edwards, R.L., Fairbanks, R.G., Friedrich, M., Guilderson, T.P., Hogg, A.G., Hughen, K.A., Kromer, B., McCormac, G., Manning, S., Ramsey, C.B., Reimer, R.W., Remmele, S., Southon, J.R., Stuiver, M., Talamo, S., Taylor, F.W., van der Plicht, J., Weyhenmeyer, C.E., 2004. IntCal04 terrestrial radiocarbon age calibration, 0–26 cal kyr BP. *Radiocarbon* 46, 1029–1058.
- Scott, E.M., 2003. The fourth international radiocarbon intercomparison (FIRI). *Radiocarbon* 2, 135–150.
- Shore, J.S., Bartley, D.D., Harkness, D.D., 1995. Problems encountered with the C-14 dating of peat. *Quaternary Science Reviews* 14, 373–383.
- Steier, P., Rom, W., 2000. The use of Bayesian statistics for C-14 dates of chronologically ordered samples: a critical analysis. *Radiocarbon* 42, 183–198.
- Telford, R.J., Heegaard, E., Birks, H.J.B., 2004. All age–depth models are wrong: but how badly? *Quaternary Science Reviews* 23, 1–5.
- van der Plicht, J., van Geel, B., Bohncke, S.J.P., Bos, J.A.A., Blaauw, M., Speranza, A.O.M., Muscheler, R., Björck, S., 2004. The Preboreal climate reversal and a subsequent solar-forced climate shift. *Journal of Quaternary Science* 19, 263–269.
- Walker, M.J.C., Coope, G.R., Sheldrick, C., Turney, C.S.M., Lowe, J.J., Blockley, S.P.E., Harkness, D.D., 2003. Devensian Lateglacial environmental changes in Britain: a multi-proxy record from Llanilid, South Wales, UK. *Quaternary Science Reviews* 22, 475–520.
- Wohlfarth, B., Skog, G., Possnert, G., Holmquist, B., 1998. Pitfalls in the AMS radiocarbon-dating of terrestrial macrofossils. *Journal of Quaternary Science* 13, 137–145.